DOCUMENT RESUME

ED 200 899 CS 005 984

AUTHOR Hayford, Paul D.
TITLE A Basic Program Analysis and Monitoring System in Reading: Alternative Monitoring Devices.
PUB DATE Mar 79
NOTE 22p.; Paper presented at the Annual Meeting of the American Educational Research Association (San Francisco, CA, April 8-12, 1979).

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Elementary Education; *Program Evaluation; Reading Achievement; *Reading Comprehension; Reading Diagnosis; *Reading Programs; *Reading Research; *Reading Tests; Test Construction; Test Items; *Test Validity
IDENTIFIERS *Program Analysis and Monitoring in Reading

ABSTRACT

The Program Analysis and Monitoring in Reading (PAM) package helps teachers and administrators anlayze, understand, and improve school reading programs by providing them with monitor reports and program reports. Both types of reports include results of student performance on tests of reading comprehension--the monitor report provides detailed information (two types of scores, passage scores and a monitor score) for each student in a reading class; the program report lists more general information, such as the average class monitor scores for each test given in a school year. During its two years of development, the PAM package has used three types of multiple choice item forms to assess literal comprehension: a modified close measure, traditional WH- detail verbatim items, and modified paraphrase items. The modified paraphrase items follow the WH- detail format while substituting, whenever possible, synonymous words or phrases for the content words in the passage sentence. The PAM package contains 24 achievement monitors--4 parallel forms at each of six levels of difficulty. Because of this, the test form a student takes is tailored to his or her ability level and he or she never takes the same form twice in a year. (Appendixes contain information on interpretation of the scores and provide rules for constructing paraphrase items.) (MKM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED200899

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)
This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.
Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

A Basic Program Analysis and Monitoring System in Reading: Alternative Monitoring Devices*

Paul D. Hayford

March 1979

Bureau of School and Cultural Research
New York State Education Department
Albany, New York 12234

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Paul D. Hayford

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

CS005988

*A paper presented at the annual meeting of the American Educational Research Association, San Francisco, California, April 8-12, 1979.

In their presentations on PAM--Program Analysis and Monitoring in Reading--Steven Kidder and Robert Ambrosino have described how the PAM Package helps administrators and teachers analyze, understand, and improve school reading programs by providing them with Monitor Reports and Program Reports. The utility of both types of reports depends on the achievement monitors used to measure pupil progress in reading. Both reports include results of pupil performance on tests of reading comprehension. The Monitor Report provides very detailed information--two types of scores, passage scores and a Monitor Score--for each pupil in a reading class. The passage scores are the percentage of items correct on each test passage. The Monitor Score is a conservative estimate of the most difficult or complex prose the pupil can literally comprehend. It is based on the pupil's pattern of performance on the test passages. The Program Report lists more general information--average class monitor scores for each test administration during a school year.

The properties of the Achievement Monitors used in the PAM Package permit estimates of progress in literal comprehension for individuals and groups both throughout a school year and from year to year. These properties are the validity of the tests as measures of literal comprehension, the ranking of test passages on a scale of difficulty from 1 to 26, sensitivity to gains over periods as short as ten weeks, and a high ceiling preventing all but the most extraordinary elementary-school pupils from topping out.

During its two years of development, the PAM Package has used three types of multiple-choice item forms to assess literal comprehension. In 1977-78, PAM used a modified cloze measure, the Multiple-Choice Cloze, in its achievement monitors. In 1978-79 PAM has switched to two other types of items, traditional WH- detail verbatim items and modified paraphrase items. The modified paraphrase items follow the WH- detail format while substituting wherever possible synonymous words or phrases for the content

words (nouns, verbs, adjectives, and adverbs) in the passage sentence. The three item forms which have been used in the PAM achievement monitors are rule-based. All three have been validated in studies reported elsewhere (O'Reilly & Streeter, 1977; Kidder, Hayford, & Salter, 1977; Hayford & Salter, 1978), and all exist in large batteries of passages scaled by difficulty through the use of readability formulas.[1] (Illustrations of each item type appear in Figures 1, 2, and 3.)

Time does not permit a thorough description of all three types of items. I will try here to focus my discussion on the paraphrase item form, which seems to hold the most promise for PAM achievement monitoring. During 1978-79, the PAM Achievement Monitors have comprised 24 test forms with a total of 720 items, 90 percent paraphrase items and 10 percent WH- verbatim items. PAM switched to paraphrase items because of their superior theoretical justification as measures of literal comprehension (Anderson, 1972). Though studies involving the Multiple-Choice Cloze and the verbatim WH- detail item have yielded very high reliability and validity coefficients,[2] both these item forms share the weakness of being susceptible to test-wiseness or test-taking skills. Orthographic matching, for example, can "solve" a WH- verbatim item, and general grammatical and semantic skills can be employed to get right answers on Multiple-Choice Cloze items in the absence of understanding of specific context. It should be noted that the verbatim

---

[1] Test batteries include 1,725 Multiple-Choice Cloze exercises, 300 passages with 8 WH- verbatim items each, and 124 passages with 6 paraphrase items each.

[2] For example, the Multiple-Choice Cloze and the WH- verbatim items have correlated in the high .60's and the .70's with the California Achievement Test in grades 1-9, and the Multiple-Choice Cloze has correlated in the high .70's and low .80's with the Gates-MacGinitie and the Stanford Achievement Test in a highly generalizable sample in grades 3, 6, and 9. (Incidentally, the paraphrase item form had similar correlations with the Gates-MacGinitie and with the Multiple-Choice Cloze.) See O'Reilly & Streeter, 1977, and Kidder, Hayford, & Salter, 1977.

Though Loki, the fire god, was ____11____ and ready-witted, his ____12____ was really evil. He was, indeed, the ____13____ of most of the ____14____ which befell the gods. He was ____15____ in trouble, yet often ____16____ because the gods valued his ____17____. It was he who ____18____ ways out of ____19____ for them, so that for a ____20____ time they felt that they could not do without him.

(11)
1. handsome
2. alcoholic
3. fragrant
4. squeaky
5. fundamental

(12)
1. pound
2. emblem
3. awl
4. nature
5. uniform

(13)
1. salvage
2. cause
3. stress
4. sphere
5. protection

(14)
1. applications
2. nominations
3. scratches
4. misfortunes
5. mosquitoes

(15)
1. scarcely
2. precisely
3. strangely
4. immensely
5. constantly

(16)
1. forgiven
2. postponed
3. quadrupled
4. puzzled
5. admitted

(17)
1. admittance
2. discomfort
3. cleverness
4. propeller
5. plasma

(18)
1. silenced
2. affirmed
3. supported
4. froze
5. found

(19)
1. difficulty
2. suspension
3. conscience
4. intermission
5. desertion

(20)
1. swampy
2. radiant
3. long
4. leaky
5. crisp

Figure 1. Multiple-Choice Cloze Exercise.

One warm day in September, Roy went for a boat ride. He went with his teacher and the class. They went on the big lake by their school. It was cool on the water. The wind was high. It made waves on the lake. All the children had fun that day. They enjoyed being near the water on such a warm day.

14. When did Roy go for a boat ride?

    A. in September
    B. in December
    C. on Saturday

15. Who did Roy go with?

    A. his brother and his sister
    B. his mother and his father
    C. his teacher and the class

16. What was high?

    A. the water
    B. the wind
    C. the wave

17. Who had fun that day?

    A. all the children
    B. the boys
    C. all the teachers

18. Where did the children enjoy being?

    A. near the water
    B. in the classroom
    C. near the school

Figure 2. PAM Achievement Monitor passage with WH- verbatim items.

I went into another room, where the walls and ceiling were all hung round with cobwebs, except a narrow passage for the artist to go in and out. At my entrance he called aloud to me not to disturb nis webs. He lamented the fatal mistake the world had been so long in of using silk-worms, while we had such plenty of domestic insects, who infinitely excelled the former, because they understood how to weave as well as spin. And he proposed farther, that by employing spiders the charge of dyeing silks would be wholly saved, whereof I was fully convinced when he showed me a vast number of flies most beautifully colored, wherewith he fed his spiders, assuring me that the webs would take on their color. And as he had them of all hues, he hoped to fit every body's fancy, as soon as he could find proper food for the flies, of certain gums, oils, and other sticky matter, to give a strength and consistence to the threads.

13. What did I enter?

A. another cave
B. another passage
C. another chamber
D. another building

14. What were the walls and ceiling completely covered with?

A. dust balls
B. spiderwebs
C. fishnets
D. silk cloth

15. What was the slender corridor for?

A. the introduction of the spiders
B. the artist's entrance and exit
C. the artist to observe and experiment
D. the beautifully colored flies

Figure 3. PAM Achievement Monitor passage with paraphrase items.

16. When did the artist cry out to me not to upset his spiderwebs?

A. when I left
B. when I came in
C. when I was fully convinced
D. when I had plenty of domestic insects

17. What did the artist deplore?

A. the excessive numbers of spiders in the world
B. the fatal mistake of using flies as food for silkworms
C. the mistake the world had been so long in of using weavers
D. the world's longstanding, deadly error of employing silkworms

18. What would be totally avoided by the use of spiders?

A. the cost of coloring silk cloth
B. the expense of cultivating silkworms
C. the charge of dyeing flies
D. the cost of killing insects

Figure 3. (Continued)

items only appear on test forms containing passages at the lowest difficulty levels--1 to 6; since only pupils just beginning to read respond to these items, there seems little danger of their responses being contaminated by test-wiseness.

The development of the test items used in PAM achievement monitoring has stressed validity, especially construct validity. This seems often to have been neglected in the development of achievement tests in reading, perhaps in the belief that no theoretical rationale is necessary for items which do, after all, involve responding to the printed word. Literal comprehension has been emphasized because it is the goal, the major objective, of elementary reading programs. All the detailed, minute instructional objectives of elementary reading programs are but means to the end of literal comprehension: understanding the explicit meaning of the written word. Further, literal comprehension is also the foundation for all other comprehension, whether it be literary prose, poetry, or the tortured polemics of the psycholinguists. Concern for the validity of the PAM achievement monitors has been warranted; the reports provided to teachers and administrators would have little practical utility for reading programs if the information they contained were based on tests which measured something other than or in addition to reading comprehension.

The PAM Package contains 24 achievement monitors, 4 parallel forms at each of six levels overlapping in difficulty (Appendix A). Because the passages on each form are scaled by difficulty and there are 6 levels of achievement monitors, the test form a pupil takes is tailored to his level of ability and he never takes the same form twice in a year. It is conceivable, for example, that a large heterogeneous fifth- or sixth-grade class could see pupils taking all 24 of the achievement monitors at a given test adminis-

tration. To be able to follow pupil progress from grade 2 through grade 6 or higher on the same scale of prose difficulty requires accurate scaling of the passages on the achievement monitors. In most cases, pupil performance on the test forms has confirmed the accuracy of the readability scores used to scale the passages. But there are several cases of reversals--that is, where groups perform better on a passage of supposedly higher difficulty than on a lower-difficulty passage. Practically speaking, for an individual pupil this occasional problem of passage scaling is taken into consideration by the algorithm used to assign a Monitor Score. If a pupil's performance is erratic (whether through his own inconsistency or through a problem in passage scaling), his score is adjusted to avoid inflation. (The three basic rules for Monitor Score assignment are explained in Appendix A.) I will return later to some recommendations for improving passage scaling.

A distinct advantage of the PAM achievement monitors is their sensitivity to change over ten-week periods. If group performances did not show change over short periods, the achievement monitors would lose much of their utility for the teacher, who in most cases already has access to information based on wide-interval testing. In 1977-78, when the Multiple-Choice Cloze was used in PAM, grade 2 pupils averaged increases of .78 points on the 1 to 26 scale of difficulty over ten-week intervals, and grade 5 pupils averaged .42 points. This year, over three test administrations using the paraphrase item form, scores for pupils in grades 2, 3, 5, and 6 have averaged the following increases, respectively, over ten-week intervals: .49, .41, .28, and .11. These increases have occurred despite a significant departure from last year's testing procedures. Last year a screening test was initially administered which placed pupils rather accurately for the first achievement monitor administration. This year the screening test was not administered,

the first test administration thus involving some mislevelling of pupils and apparently deflating increases from TA1 to TA 2.

A final advantage of the paraphrase items used in PAM is their high ceiling. For example, 8 pupils out of some 4500 have topped out on the highest level (level 6) PAM monitors. These pupils were sixth-graders. Last year, in contrast, with the cloze items, at least five times as many pupils (fifth-graders) from a sample half as large topped out at level 6. What this suggests is the potential of extending PAM into junior high school, and even into senior high for students with reading problems. Indeed, one administrator is planning next year to follow his seventh-grade students who are in compensatory programs.

The experience of developing PAM and its paraphrase items has resulted in the acquisition of practical information which may be worth conveying. The observations which follow are related, respectively, to item writing, validity and reliability, and passage scaling.

The advantages of writing items to rules (Appendix B) are standardization and elimination of subjectivity and idiosyncratic intepretation. But writing paraphrase items is not simple. It involves three problems. One is that not all sentences, clauses, phrases, etc. can be paraphrased. Not all sentences will yield items (and awkwardness and barbarousness should at all costs be avoided). A second is that it is hard to control vocabulary difficulty when paraphrasing. We have tried to do this by using graded vocabulary lists, but it will not always work. Hard sentences usually result in hard paraphrases. The third problem is related to the artfulness involved in item writing. In this case the art is not involved in concocting items, it is involved in identifying sentences which will yield items (and having the ability quickly to recognize a sentence destined to remain fruit-

less). It should also be noted that a staff endeavoring to produce large batteries of such items (or any similar items) must include someone with high-level editorial skills.

The three item types discussed in this paper have been employed in large-scale validation studies. Correlational studies have discovered high reliability and validity coefficients. But the relevance of such traditional estimates of reliability and validity is questionable for tests which are used as the PAM monitors are. For instance, in the validation studies samples were generalizable and randomly selected and tests were constructed for the sake of yielding maximum variance. But in practice both the range of passage difficulty in test forms and the range of pupil ability are greatly restricted. Also, total test scores are not used in PAM, and the monitor scores which are used involve another restriction affecting variance. The range of possible Monitor Scores on a given test form is about 10 points, whereas a total test score range could be 30 points.

Four suggestions for improving scaling are proposed. (1) Revise items further. (2) Rescale passages using Rasch analysis to produce a new, Rasch scale. (3) Rescale passages using empirical difficulty (a bit of a problem here, because the difficulty levels would come into question). (4) Produce a new battery of passages with (a) a simpler scale, say, 1 to 13, and with (b) clearer gaps (in readability scores) between passages on the scale.

References

Anderson, R.C. How to construct achievement tests to assess comprehension. Review of Educational Research, 1972, 42 (2), 145-170.

Hayford, P.D., & Salter, R. Rule-based measures of literal comprehension. Paper presented at the annual meeting of the American Educational Research Association, Toronto, March 1978.

Kidder, S.J., Hayford, P.D., & Salter, R. Construct validation of Multiple-Choice Cloze Exercises. Draft report, Bureau of School and Cultural Research, New York State Education Department, Albany, NY, 1977.

O'Reilly, R.P., & Streeter, R.E. Report on the development and validation of a system for measuring literal comprehension in a multiple-choice cloze format: Preliminary factor analytic results. Journal of Reading Behavior, 1977, 9 (1), 45-69.

APPENDIX A

The University of the State of New York
THE STATE EDUCATION DEPARTMENT
Bureau of School and Cultural Research
Albany, New York 12234

PAM--PROGRAM ANALYSIS AND MONITORING IN READING

Monitor Reports and Monitor Scores

Pupil performance on the Literal Comprehension Achievement Monitors used in the PAR Project is reported by reading class after each test administration. The information in the several columns of the computerized Monitor Report and the determination of monitor scores are described below.

Monitor Report Columns

PUPIL--All pupils who are or have been members of the reading class are listed here. Pupils no longer in the class remain on the list because their previous monitor scores contribute to class averages.

TEST FORM--This column shows which test form the pupil took. There are 24 Achievement Monitors, four each at six levels with overlapping difficulty ranges. A double asterisk here indicates that the pupil did not take the test.

PASSAGE SCORE--The six columns under this heading show the pupil's score on each passage of his achievement monitor. Only the level 1 forms (11, 12, 13, and 14) have six passages; all others have five. The passing score for a passage is 75%. A series of -1's in the passage score columns indicates absence from the test or incomplete test data.

NEXT TA--The +'s and -'s in this column mark pupils who are to be given higher- or lower-level test forms at the next test administration. The general rule is that a pupil will take a lower level form if he failed all passages and a higher level form if he passed all passages or all but one passage.

MONITOR SCORE--The Monitor Score, based on passage performance, is a difficulty level from 1 to 26, indicative of the readability of the material the pupil can literally comprehend. The higher the monitor score, the more difficult the material. (How monitor scores are determined from passage scores is described in detail below.) Double asterisks in Monitor Score columns indicate absence.

The Average Monitor Score for the class is the arithmetic mean of the individual monitor scores.

The Monitor Report gives Monitor Scores and Average Monitor Scores for the most recent test administration and all preceding administrations, allowing progress in literal comprehension to be followed over time.

Determination of Monitor Scores

As noted above, the monitor score is determined by the pupil's passage scores. If a pupil performs consistently, passing each successive passage on his achievement monitor until the passages become too difficult for him, his monitor score will be the difficulty level of the last passage on which he scored 75% or better. If, however, the pupil fluctuates, passing some passages and failing others, his monitor score will be the difficulty level of the second most difficult passage passed. If the pupil passes only one passage or fails all of the passages, his monitor score will be the difficulty level of the first (easiest) passage.

To demonstrate, passage and monitor scores for four pupils taking form 41 appear below.

Pupil	Passage Scores (Difficulty level in parentheses) 1(11)	2(13)	3(15)	4(17)	5(19)	6(*)	Monitor Score
Susan Jones	100	100	83	83	67	0	17
Tom Smith	100	67	50	83	83	0	17
Terri Watson	67	83	50	83	67	0	13
Kim Young	50	67	33	83	50	0	11

*No passage 6 on Form 41.

Susan Jones was consistent. She passed the first four passages and failed the last. Her monitor score is therefore 17, the difficulty level of the fourth passage.

Tom Smith was inconsistent. He passed the first passage, failed the next two, and passed the fourth and fifth. His monitor score is also 17, the difficulty level of the second most difficult passage on which he had a passing score.

Terri Watson failed, passed, failed, and passed again before failing the last passage. Her monitor score is 13.

Kim Young had a passing score on only one passage, the fourth. Her monitor score is 11.

The attached chart shows the difficulty levels of the passages on each of the test forms. It should be used in relating monitor scores to passage scores and analyzing pupil performances.

Literal Comprehension Achievement Monitors
Difficulty Levels of Passages by Form

Level 1				
Passage	Form 11	Form 12	Form 13	Form 14
1	1	1	1	1
2	2	2	2	2
3	3	3	3	3
4	4	4	4	4
5	5	6	5	6
6	7	8	7	8
Level 2				
	Form 21	Form 22	Form 23	Form 24
1	4	5	4	5
2	6	7	6	7
3	8	9	8	9
4	10	11	10	11
5	12	13	12	13
Level 3				
	Form 31	Form 32	Form 33	Form 34
1	7	8	7	8
2	9	10	9	10
3	11	12	11	12
4	13	14	13	14
5	15	16	15	16
Level 4				
	Form 41	Form 42	Form 43	Form 44
1	10	11	10	11
2	12	13	12	13
3	14	15	14	15
4	16	17	16	17
5	18	19	18	19
Level 5				
	Form 51	Form 52	Form 53	Form 54
1	13	14	13	14
2	15	16	15	16
3	17	18	17	18
4	19	20	19	20
5	21	22	21	22
Level 6				
	Form 61	Form 62	Form 63	Form 64
1	17	18	17	18
2	19	20	19	20
3	21	22	21	22
4	23	24	23	24
5	25	26	25	26

APPENDIX B

RULES FOR CONSTRUCTING PARAPHRASE

ITEMS FOR PAM ACHIEVEMENT MONITORS

I. Passage Selection

A. Determine range of difficulty for test forms.

1. Identify each difficulty level in the Reading/Literature MCC Exercises from which passages will be drawn.

2. Draw randomly the requisite number of exercises at each difficulty level.

3. Replace deleted words in blanks in each MCC exercise drawn.

II. Paraphrasing Selected Exercise Passage *

A. Number each sentence in every exercise passage.

1. In passages with compound sentences, number each main clause.

2. In passages with complex sentences, number each main clause, subordinate clause, and long modifying phrase.[1]

B. Paraphrase[2] each numbered sentence or clause.

1. If possible, replace all substantive words (nouns, verbs, modifiers[3]) with synonyms[4] (i.e., equivalent words or phrases).

a. Consult when necessary a dictionary, thesaurus, or dictionary of synonyms.

b. Consult other relevant reference words as necessary.

2. Proper nouns and pronouns often cannot be paraphrased.

3. Auxiliary verbs and the verb <u>to be</u> cannot always be paraphrased.

4. If possible, paraphrase vocabulary should not exceed the vocabulary level of the passage (as determined by difficulty level).

a. Consult Harris and Jacobson, 1972, when necessary.

b. Consult Carroll, Davies, and Richman, 1971, when necessary.

5. Retain meaning of original sentence (i.e., vocabulary and syntax of paraphrase should not involve significant alteration of the literal meaning of the original sentence).

---

* Rules for paraphrasing are based on Anderson's (1972) definition of paraphrase.

C. Flexibility in the writing of paraphrases is illustrated below:

1. A paraphrase does not have to have the exact number of words as the original sentence; it may be slightly longer or shorter.

2. Syntax may be altered in various ways.

a. Order of clauses or phrases may be changed as long as literal meaning is retained.

b. Voice of verbs may be changed (e.g., active to passive).

c. Phrases may replace single words (and vice versa).

III. Writing Items for Paraphrased Passages[5]

A. Write WH-detail items on each paraphrased sentence, clause, or phrase. Adhere as much as possible to the following rules:

1. Write clear, concise questions in colloquial English, changing the wording of the paraphrase as little as possible. (Exception: replace pronouns with their referents.)

2. Begin each question with the appropriate detail word (e.g., how, what, when, where, etc.).

3. Avoid writing inferential WH-detail items (e.g., do not write a "why" item unless the causal relationship is either explicit or clearly implied in the text).

4. Write as many WH-detail items as possible for each paraphrase.

5. Try to write as least two WH-detail items for each paraphrase Note: Requirement for test forms was six WH-detail items/ passage. Passages are very short (50-80 words).[6]

B. Write three distractors for each item (i.e., four responses, including distractors and correct response).

1. Write only grammatically and semantically plausible distractors.

2. Write parallel distractors when possible.

3. Write distractors that closely match the correct response in number of words.

4. Avoid writing response arrays in which the correct response characteristically stands out because of its brevity, length, or syntax.

5. Write no distractors that could be correct in the context of the passage.

6. Write distractors that are appropriate to the difficulty level of the passage (see II. B. 4, above).

IV. Problems and Responses

A. Paraphrases

1. Not every sentence yields an adequate paraphrase, For example, vocabulary levels, uniqueness of vocabulary or structure, and other factors may make paraphrasing difficult.

2. When sentences which cannot be acceptably paraphrased result in passages which do not yield the requisite number of items, select another passage randomly from the relevant difficulty level.[7]

B. Items

1. When item stems contain substantive words verbatim from the passage, make sure correct response is not verbatim (i.e., do not write verbatim WH- detail items).

2. When a correct response is verbatim, make sure that some distractors are also verbatim to diminish the possibility of orthographic matching.

3. When a correct response is partially verbatim (e.g., this occurs occasionally in longer responses), make sure at least one distractor contains the verbatim element which appears in the correct response (to diminish orthographic matching).

Footnotes

[1]Extracted from context, subordinate clauses and some phrases may be paraphrased as main clauses or sentences. Example: "But even [a liar's invention] , being an empty thing that offers no hold . . ." is paraphrased as "a prevaricator's fiction is a vacuous thing that provides no handle" for a wh-item as follows: "What kind of thing is a prevaricator's fiction?"

[2]Note: An alternate version of a sentence, clause, or phrase which "means" what another sentence, clause, or phrase "means" is not necessarily a paraphrase according to the rules here presented. Saying a thing in another way is not always equivalent to paraphrasing by these rules.

Such a situation occurs on occasion when a reviewer is dissatisfied with an item stem (or stem plus response) and rewrites the item to make it sound better or to avoid heaviness, awkwardness, wordiness, etc.--but without first writing a new paraphrase or without taking the original paraphrase into consideration. The rewritten item, considered out of context, will often sound or look better, but it will often no longer be an item based on an acceptable paraphrase.

A similar problem arises when an item is rewritten but is no longer a WH-detail item.

[3]Modifiers include adjectives and adverbs, not articles or determiners.

[4]Superordinate terms are not necessarily acceptable synonyms (e.g., dog is not necessarily an acceptable synonym for Siberian wolf-hound).

[5] See Rules for Constructing WH-Detail Items, on file with BSCR.

[6]Average number of WH-detail items written for each passage was more than ten, of which six were selected. Criteria for selection were quality (e.g., absence of awkwardness and turgidity) and freedom from mutual cueing, defined as a stem giving away a response to another stem. In the following, for example, stem A cues the answer to stem B: "A. When did the fuel drums burst into flame?" "B. What burst into flame?"

[7]Fewer than ten per cent of the passages from the original sample had to be replaced.